

## Machine Readable Archives



ISSN 0821-3658

Publications.

Vol. 1 – No. 3 Fall 1983

## Reference System

As part of its mandate, the Machine Readable Archives provides as much descriptive information about its holdings as possible. As machine readable records are not in a visual form, it is important to provide multiple access points to the information they contain. The division has developed a "Reference System" with several levels of description. Although not as yet automated, each level provides more descriptive detail about the contents of the data files.

The Reference System contains four main elements: the automated inventory, the accession register, the card catalogues, and published finding aids. The system was designed to ensure the control of the informational content of the holdings and to provide a sound base from which both researchers and archivists can obtain information.

A brief description of the contents and uses of each component is provided.

#### Automated Inventory of Canadian Machine Readable Data Files

The automated inventory provides access to information on EDP record holdings in the federal government and on nationally-significant data that exist in the private sector. Information on the records has been obtained from federal government inventories (1976 and 1981), publications and research undertaken by divisional archivists. Although far from complete, the automated inventory does contain minimal descriptions on over 5,000 files. The major descriptive elements are: file title; creating agency; file location; storage medium and volume; type of file; appraisal value; acquisition decision; and content description. The description of the contents is only as complete as the information that the department or agency provided. The inventory is the only automated part of the reference system at this time. The data are accessible through the use of the MINISIS information management package. Access is presently limited to the staff of the division. As part of its public service function, the Documentation and Public Service Section will help researchers locate files that may be of interest.

## Accession Register

As EDP records are identified and appraised some of them are acquired by the division for their long term value. The Acces-

sion Register records the following: title; transferring department or organization; acquisition date; responsibility centre; type of file; and a brief description of the contents. A preliminary catalogue record is developed and entered into the system at the time of accessioning. The data file is thus recorded into the holdings, although processing of the data has not started.

#### Card Catalogues

Once the data files have been processed and documented, a full range of finding aids are developed. The division provides bibliographic information on its holdings using the Anglo-American Cataloguing Rules for machine readable data files. The main elements described are: principal investigator; title; responsibility statement; producer, date and place of production; distributor, place and date of distribution; size of the file; type of documentation; summary description of the contents; access points; and any particular detail that makes the data file unique. The data files are accessible by title, principal investigator/organization, and by subject. It is currently a manual system, although automation of the system and the preparation of a MARC record for distribution to other archives is presently under consideration.

Two other card catalogues also exist: the Variable Index and the Record Group Index. The Variable Index provides a detailed list of information elements in the data file and can be used for quick reference to specific variables contained in the file. The Record Group Index organizes the data files by provenance and is used to relate the holdings of the MRA to those of the Federal Archives Division. The latter division is responsible for the historically-valuable textual records created by federal government departments and agencies.

## **Published Finding Aids**

A number of published finding aids describing the divisional holdings are produced for distribution to researchers, data archives and data libraries. The major publication is the *Catalogue of Holdings*, which provides descriptive information on all files that have been processed. This publication is indexed by title, principal investigator/organization, and subject. Other smaller but more specialized publications are produced. These pamphlets are subject-oriented and are developed for specific research groups. Their preparation is dependent upon the acquisition of a number of files in a specific field.

Two current publications are the *Alcohol*, *Drug and Tobacco Use Data Files* and *Recreation and Leisure Data Files*.

There is often a time lapse between the acquisition of a data file, the processing of the file and the appearance of a published description. In order to make researchers aware of new acquisitions, a section of this *Bulletin* is used to provide information on recently accessioned data files.

Although the divisional Reference System is mostly a manual one, it is hoped that full automation of the components will be available over the next two years. Not only will an automated system provide quicker access, but it will also enable the publications to be prepared more rapidly.

## Citation of Machine Readable Data Files

A Machine Readable Data File (MRDF) is defined as a body of information encoded and formatted in such a way that it requires the use of a data processing machine or computer to be properly interpreted. Such files are being used increasingly for secondary analysis. Over the past few years, standards for their proper citation have been developed and are being used. In the following paragraphs the elements required in the citation will be identified, the format of the citation explained, and examples of citations provided.

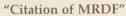
#### Elements Required in the Bibliographic Citation of Machine Readable Data Files

Authorship: The person or institution primarily responsible for the intellectual content of the MRDF such as the principal investigator, the department or agency, the project director.

*Title:* The full descriptive title of the MRDF including dates. No acronyms are acceptable as full titles.

Material Designator: The statement [machine readable data file] enclosed in square brackets must be given. The material designation is required to indicate that the information is in computerized form.

Statement of Authorship: The indication of the relationship of the work to the person(s) or corporate body named as principal investigator or to other significant parties, such as the



cont'd from page 1

department, funding agency or sponsoring organization.

Edition Statement: The indication of which MRDF is being described. A change in edition occurs if major additions or deletions of the original MRDF are made or if the variables within the file are altered; major reformatting or major changes in the number of observations or logical records; and changes in the programming statements or language.

Producer Statement: The person or corporate body with the financial and/or administrative responsibility for the creation of the MRDF. The statement includes the name of the producer, as well as the place and date of production.

Distributor Statement: The person or organization that has been designated by the author or producer to reproduce copies of the MRDF being cited. The statement includes the place and date of distribution, as well as the name of the distributor.

#### Use of Square Brackets for Producer and Distributor Statement

Square brackets are used around the words [producer] and [distributor] in the producer and distributor statements if confusion exists as to the producer and distributor. The use of the square brackets is to clarify the role of each. This is an optional convention.

#### **Format**

Surname, Christian names of principal investigator. Title [general material designator]. Responsibility statements. Edition statement. Place of production: name of producer, date of production. Place of distribution: name of distributor, date of distribution.

#### Examples

- 1. Pay Research Bureau. Hospital Survey of Nurses, Social Workers, and Other Hospital Classes, 1974 [machine readable date file]. 1st. ed. Ottawa: Public Service Staff Relations Board, 1974. Ottawa: Public Archives Canada, Machine Readable Archives, 1977.
- 2. Hayes, James. Citizen Participation in Non-Work Activities [machine readable data file]. 1st. ed. Ottawa: Secretary of State, Citizenship Branch [producer], 1975. Ottawa: Public Archives Canada, Machine Readable Archives [distributor], 1977.
- 3. Survey of Consumer Attitudes and Behaviour, 1977 [machine readable data file]. Conducted by University of Michigan for Industry, Trade and Commerce/Canadian Government Office of Tourism. Ottawa: Industry, Trade and Commerce/Canadian Government

Office of Tourism [producer], 1977. Ottawa: Public Archives Canada, Machine Readable Archives Division [distributor], 1979.

## **Documentation Manuals**

Documentation manuals, or codebooks as they are often called, are the basic finding aid for machine readable data files. As such they require care in their preparation to ensure the usefulness of the data in secondary research. Although all data libraries and archives have defined the elements in the preparation of documentation manuals, the format is usually defined by the particular institution. It has always been hoped that archival considerations could be built into the original research project in processing the data file. The preparation of good documentation also provides the principal investigator with an organized document from which to work.

The MRA prepares a documentation manual for each processed file following internal standards. The divisional standards are based on information from other archival institutions and associations of data users, as well as internal requirements to retain administrative information specific to the Public Archives. A complete documentation manual contains administrative information (transfer agreements, restrictions); the codebook; and printouts of programs used to process the data (dumps of labels and records and frequency runs). As the administrative information that is kept is specific to the institutions preparing the documentation, only the guidelines for the preparation of the codebook will be described.

#### Codebooks

A codebook should provide sufficient information to the user so that he/she can use the data file without help from the distributing agency. The elements that should be contained are described below. The amount of detailed description is very dependent upon the information provided by the creator of the data file.

Bibliographic Information: Elements describing the bibliographic details of the data file are contained on the title page. These elements include a descriptive title including dates; the manual(s) of the principal investigator; the manual(s) of the sponsoring organization; the name of the institution responsible for the data collection; date and place of data production; the name of the distributor; date and place of distribution; and the edition statement.

Purpose of the Machine Readable Data File: A description outlining the reason(s) for undertaking the collection of the data.

Study Description: A detailed description of the contents of the data file. Methodology: Information on how the data were collected and what type of data collection instrument was used.

Survey Design: A full description of the survey design, the origin of the sample, the sample size, the percentage of returns, weights and weighting procedures, as well as geographic coverage are included.

Processing Information: The verification process is described. Any particular details on problems encountered, errors found, and specific checks that were made by the archivist are written up.

Related Publications and Uses of the Data: This section lists all reports produced as a result of the data collection. An attempt is made to obtain information from users of the data file in order to provide examples on how the data can be used.

Variables/Values: A detailed description of all the variables, their location, their type and their values is provided. Any errors found in the data are documented. All variables are referenced to the source document.

Technical Information: The technical specification of the data file such as volume serial number, format, label, track, density, mode, parity, blocking factor, etc., are recorded. This information forms part of the documentation required.

These are the basic requirements for documentation prepared by the division. The amount of detail does vary from one file to another. The variance depends on how much detail the creator provided. A minimum of detail is available for all of the holdings as documentation is one of the elements studied in the appraisal of the records.

## **New Acquisitions**

The division has acquired a series of files from Environment Canada titled, The National Water Quality Data System. The files contain data on the quality of water since 1960

Forty-four surveys of National Parks were acquired from the Quebec Region of Parks Canada. These surveys were undertaken between 1977 and 1982 and are specific to National Parks located in Quebec.

But du fichier ordinolingue : Description expliquant pourquoi on a rassemblé des données.

Description de l'étude : Description détaillée du contenu du fichier. Méthodologie : Renseignements sur la façon

Methodologie: Kenseignements sur la taçon dont les données ont été saisies et sur les outils utilisés.

Conception de l'enquête; Détails sur les principes sous-tendant l'enquête, le choix et l'importance de l'échantillon, le pourcentage de réponses, le poids et les procédés connexes, et la portée géographique.

Renseignements sur le traitement: Description du processus de vérification. Y sont consignés fous les détails au sujet des problèmes rencontrés, des erreurs commises et des vérifications périodiques effectuées par l'archiviste.

Publications connexes et utilité des données: Liste de tous les rapports produits suite à la saisie des données. On essaie d'obtenir des renseisnemnts auprès des usagers du fichier pour voir comment les données peuvent être utilisées

Variables et valeurs: Description détaillée de toutes les variables: emplacement, type et valeurs. Toutes les erreurs sont documentées et les variables contiennent toutes des renvois au document de base.

Renseignements techniques: Sont décrites les normes techniques du fichier de données telles que le numéro de série du volume, le format, l'étiquette, les pistes, la densité, le mode, la parité, le facteur de groupage. Toutes ces informations font partie de la documentation requise.

Voilà les éléments fondamentaux de la documentation préparée par la division. La quantité de renseignements varie cependant d'un fichier à un autre compte tenu des informations fournies par l'auteur du fichier. Toutefois, il y a un minimum d'informations pour chaque fichier car la documentation est un des éléments compris dans l'évaluation des dossiers.

## Nouvelles acquisitions

La division a acquis une série de fichiers du ministère de l'Environnement intitulée The Mational Water Quality Data System. Ces fichiers contiennent des données sur la qualité de l'eau depuis 1960.

Quarante-quatre enquêtes de parcs nationaux ont été acquises de la région de Québec de Parcs Canada. Ces enquêtes, menées entre 1977 et 1982, portent uniquement sur les parcs nationaux au Québec.

tawa: Industrie et Commerce, Office du tourisme du Canada [producteur], 1977. Ottawa: Archives publiques du Canada, Division des archives ordinolingues [distributeur],

## Manuels de documentation

principal tion facilite également le travail du chercheur création du fichier. Une bonne documentaprojet de recherche original au moment de la intégrer des considérations archivistiques au intéressé. On espère depuis toujours pouvoir habituellement déterminée par l'organisme documentation, la présentation matérielle est ments à incorporer dans les manuels de blissements d'archives aient défini les élédaire. Bien que les bibliothèques et les étapour qu'ils soient utiles à la recherche seconlingues. Il taut donc les préparer avec soin recherche fondamental des fichiers ordinoappelés listes de codage, sont l'instrument de Les manuels de documentation, souvent

livre de codage seront décrites. les directives s'appliquant à la préparation du nisme qui a préparé la documentation, seules nèes administratives sont propres à l'orgagrammes des tréquences). Comme les dondes labels, des enregisfrements et les proservi au traitement des données (imprimés les imprimantes des programmes qui ont transfert et restrictions), le livre de codage et formations administratives (modalités de de documentation complet contient des inpropres aux Archives publiques. Un manuel ce qui est des informations administratives aussi des desoins internes de la division pour d'utilisateurs de données; on tient compte établissements d'archives et des associations d'après les informations fournies par d'autres Les normes de la division sont élaborées de documentation pour chaque fichier traité. Archives ordinolingues préparent un manuel C'est à partir de normes internes que les

### Livres de codage

Un livre de codage doit contenir suffisamment d'informations pour que l'usager puisse utiliser le fichier de données sans l'aide du distributeur. Les éléments essentiels sont décrits ci-dessous. Plus l'auteur du fichier de données fournit d'informations, plus la description sera détaillée.

Renseignements bibliographiques: Les éléments bibliographiques sont énumérés à la page de titre. Ces éléments comprennent un titre descriptif y compris les dates, le(s) manuel(s) du chercheur principal et celui (ceux) de l'organisme de parrainage, le nom de l'établissement chargé de la saisie des données, la date et le lieu de production des données, le nom du distributeur, le lieu et la date de distribution et l'édition.

«Références»

I ageq al ab atiue

Edition: Indication du fichier ordinolingue consulté. Il y a changement d'édition si le fichier subit d'importantes modifications sont modifiées; s'il y a restructuration majeure des données ou d'importants changements dans le nombre d'observations ou de notices logiques; si des modifications sont apportees logiques; si des modifications sont apporte logiques; si des modifications sont apporte logiques; si des modifications sont apporter logiques;

Mention du producteur : Il s'agit de la personne ou de la collectivité ayant la responsabilité financière ou administrative, ou les deux, du fichier. La mention comprend le nom du production ainsi que le lieu et la date de production.

Mention du distributeur : La personne ou l'organisation que l'auteur ou le producteur a autorisée à reproduire le fichier. La mention comprend le lieu et la date de distribution et le nom du distributeur.

# Utilisation de crochets pour les mentions du producteur et du distributeur

Les mots «producteur» et «distributeur» sont placés entre crochets dans les mentions correspondantes s'il y a des doutes quant à l'identité du producteur et du distributeur. Ces crochets permettent de clarifier le rôle de chacun mais ils sont facultatifs.

#### Présentation

Nom de famille du chercheur principal suivi de ses prénoms. Titre [support]. Mention de responsabilité. Édition. Lieu de production: nom du producteux, date de production. Lieu de distribution: nom du distributeur, date de distribution.

### Fxembles

I. Bureau de recherche sur les traitements. Enquête sur les infirmières, les travailleurs sociaux et les autres travailleurs des services hospitaliers, 1974 [fichier ordinolingue]. Irê édition. Ottawa: Commission des relations de travail dans la Fonction publique, 1974. Ottawa: Archives publiques du Canada, Division des archives ordinolingues, 1977.

2. Hayes, James. Participation des citoyens aux activités récréatives [fichier ordinolingue]. Ire édition. Ottawa: Secrétariat d'État, Direction de l'enregistrement de la citoyenneté, 1975. Ottawa: Archives publiques du Canada, Division des archives ordinolingues, 1977.

3., Enquête sur les attitudes et comportements des consommateurs, 1977 [fichier ordinolingue]. Menée par l'université du Michigan pour l'Office de tourisme du Canada du ministère de l'Industrie et du Commerce. Ot-

Automne 1983



Vol. 1 - no 3

# OLLETIN

l'alcool, des drogues et du tabac et les Dossiers des ces publications sont les Dossiers sur l'usage de sant de fichiers sur un sujet donné. Deux de sont préparées que s'il y a un nombre suffigroupes de recherche particuliers. Elles ne 8595-1580 NSSI

données sur les loisirs et la récréation.

Bulletin. acquisitions seront annoncées dans ce d'un ouvrage. Par conséquent, les nouvelles données, son traitement et la publication s'écoule entre l'acquisition d'un fichier de Il arrive souvent qu'un laps de temps

publications. lèrera également la préparation des tème facilitera l'accès aux données et accèprochaines années. L'automatisation du sysrons l'automatiser en entier au cours des deux sion soit essentiellement manuel, nous espè-Bien que le système de référence de la divi-

## ordinolingues Références aux fichiers

ples à l'appui. la disposition et fournissent quelques exemments à inclure dans une rétérence, précisent Les paragraphes ci-dessous indiquent les élèété élaborées et sont maintenant utilisées. mes régissant les références à ces fichiers ont secondaires. Ces dernières années, des norplus en plus utilisés dans le cadre d'analyses interpréter le contenu. Ces fichiers sont de qu'il faut recourir à un ordinateur pour en données codées et structurées de telle façon Un fichier ordinolingue est un ensemble de

## des fichiers ordinolingues référence bibliographique Eléments à inclure dans la

l'organisme, le directeur du projet. ex. le chercheur principal, le ministère ou premier responsable du contenu du fichier, p. Auteur: La personne ou l'organisme qui est le

ne sont pas acceptés. dates de saisie des données. Les acronymes Titre: Le titre complet du fichier y compris les

Support: L'expression «fichier ordinolingue»

informatise. les données sont conservées sur un support doit figurer entre crochets pour indiquer que

parrainage. de financement ou l'organisme de significatifs tels que le ministère, l'organisme chercheur principal ou des autres éléments personne ou collectivité nommée comme Mention de responsabilité: Responsabilité de la





encore commence. même si le traitement des données n'a pas fichier de données est donc intégré aux tonds tème au moment de l'enregistrement. Le d'abord composée et introduite dans le sysdu contenu. Une notice de catalogue est tout lité, le type de fichier et une brève description la date d'acquisition, le centre de responsabitère ou l'organisme qui a transféré le fichier, tient les éléments suivants : le titre, le minisvaleur permanente. Le registre d'entrée con-

## Catalogues sur fiches

rait distribuer aux autres établissements et de préparer une notice MARC qu'on pouren train d'étudier la possibilité d'automatiser tuellement d'un système manuel, mais on est culler ou organisme) et du sujet. Il s'agit acpartir du titre, du chercheur principal (partifichier. On a accès aux fichiers de données à détail qui confère un caractère unique au sommaire du contenu, points d'accès, et tout fichier, type de documentation, description lieu de distribution, importance matérielle du et lieu de production, distributeur, date et mention de responsabilité, producteur, date vants: nom du chercheur principal, titre, lingues. Les principaux éléments sont les suianglo-américaines pour les fichiers ordinographiques à l'aide des Règles de catalogage che. La division compose ses notices bibliogamme complète d'instruments de recherdocumentés, on passe à l'élaboration d'une Une fois les fichiers de données traités et

ministères et organismes fédéraux. ments historiques traditionnels crees par les ves fédérales; celle-ci est chargée des docula division avec ceux de la Division des archiprovenance et fait correspondre les fonds de ves est une liste des fichiers d'après leur bles du fichier. L'Index des groupes d'archimet de consulter rapidement diverses variacontenus dans le fichier de données et perénumère tous les éléments d'information groupes d'archives. L'Index des variables fiches: l'Index des variables et l'Index des Il existe en outre deux autres catalogues sur d'archives.

## sənqnd Instruments de recherche

publications spécialisées s'adressant à des nisme) et sujet. Il existe également de petites chercheur principal (particulier ou orgafichiers déjà traités. Il est indexé par titre, le Catalogue des sonds qui décrit tous les thèques. Le plus important est sans contredit aux chercheurs, aux archives et aux bibliotonds de la division sont publiès et distribués Des instruments de recherche décrivant les

## Système de rétérence

səəuuop əp des informations sur le contenu des fichiers sont pas encore automatisés mais fournissent sieurs niveaux de description. Ceux-ci ne «système de référence» comprenant plud'accès. La division a donc élaboré un il est important d'en multiplier les points contrairement aux documents traditionnels, dossiers informatisés ne sont pas «visibles», mation possible sur leurs fonds. Comme les ordinolingues doivent fournir le plus d'infor-Dans le cadre de leur mandat, les Archives

et aux archivistes. fonds et d'en faciliter l'accès aux chercheurs trôler les informations contenues dans les bliés. Le système a été conçu en vue de contiches et des instruments de recherche putisé, un registre d'entrée, des catalogues sur éléments principaux : un inventaire automa-Le système de référence contient quatre

Voici une brève description du contenu et

de l'utilité de chaque élément.

## fichiers ordinolingues canadiens Inventaire automatisé des

localiser les fichiers susceptibles de les et du service au public aide les chercheurs à y avoir accès. La Section de la documentation moment, seul le personnel de la division peut nées contenues dans l'inventaire et, pour le actuelle. Le MINISIS donne accès aux donautomatisé du système de référence à l'heure l'organisme. L'inventaire est le seul élément d'informations fournies par le ministère ou cription est proportionnelle à la quantité sition et la description du contenu. Cette desvaleur estimée, la décision au sujet de l'acquitreposage et le volume, le type de fichier, la nisme créateur, l'emplacement, le mode d'ensont principalement le titre du fichier, l'orgade 5 000 fichiers. Les éléments descriptifs tient des descriptions élémentaires d'au-delà qu'incomplet, l'inventaire automatisé connées par des archivistes de la division. Bien 1981), de publications et de recherches meventaires du gouvernement tédéral (1976 et tance nationale. L'information est tirée d'inqouvees qu secteur privé qui ont une imporgues du gouvernement fédéral et sur les tormation sur les tonds d'archives ordinolin-L'inventaire automatisé donne accès à l'in-

## Kegistre d'entrée

sion acquiert certains d'entre eux pour leur formatisés sont identifiés et évalués, la divi-Au fur et à mesure que les documents in-